

I. Evans

Re-run 0
4

JCE Rec'd PCT/PTO 2 8 FEB 2002
BIOTECHNOLOGY SYSTEMS BRANCH
RECEIVED
FEB 2 8 2002
TECH CENTER 1600/2900

**RAW SEQUENCE LISTING
ERROR REPORT**

1644

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 09/807,802
Source: PCT 09
Date Processed by STIC: 5-1-01

RECEIVED
SEP 0 9 2002
TECH CENTER 1600/2900

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.
PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:
1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY
FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.

FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.
PATENTIN 2.1 e-mail help: patin21help@uspto.gov or phone 703-306-4119 (R. Wax)
PATENTIN 3.0 e-mail help: patin3help@uspto.gov or phone 703-306-4119 (R. Wax)

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE CHECKER
VERSION 3.0 PROGRAM, ACCESSIBLE THROUGH THE U.S. PATENT AND
TRADEMARK OFFICE WEBSITE. SEE BELOW:

Checker Version 3.0

The Checker Version 3.0 application is a state-of-the-art Windows based software program employing a logical and intuitive user-interface to check whether a sequence listing is in compliance with format and content rules. Checker Version 3.0 works for sequence listings generated for the original version of 37 CFR §§1.821 - 1.825 effective October 1, 1990 (old rules) and the revised version (new rules) effective July 1, 1998 as well as World Intellectual Property Organization (WIPO) Standard ST.25.
Checker Version 3.0 replaces the previous DOS-based version of Checker, and is Y2K-compliant. Checker allows public users to check sequence listings in Computer Readable form (CRF) before submitting them to the United States Patent and Trademark Office (USPTO). Use of Checker prior to filing the sequence listing is expected to result in fewer errored sequence listings, thus saving time and money.

Checker Version 3.0 can be down loaded from the USPTO website at the following address:
<http://www.uspto.gov/web/offices/pac/checker>

New Sequence Listing Error Summary

ERROR DETECTED SUGGESTED CORRECTION

SERIAL NUMBER: 09/807,802

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

- 1 ☒ **Wrapped Nucleics** The number/text at the end of each line "wrapped" down to the next line.
This may occur if your file was retrieved in a word processor after creating it.
Please adjust your right margin to .3, as this will prevent "wrapping".

- 2 ☐ **Wrapped Aminos** The amino acid number/text at the end of each line "wrapped " down to the next line.
This may occur if your file was retrieved in a word processor after creating it.
Please adjust your right margin to .3, as this will prevent "wrapping".

- 3 ☐ **Incorrect Line Length** The rules require that a line not exceed 72 characters in length. This includes spaces.

- 4 ☐ **Misaligned Amino Acid Numbering** The numbering under each 5th amino acid is misaligned. This may be caused by the use of tabs between the numbering. It is recommended to delete any tabs and use spacing between the numbers.

- 5 ☐ **Non-ASCII** This file was not saved in ASCII (DOS) text, as required by the Sequence Rules.
Please ensure your subsequent submission is saved in ASCII text so that it can be processed.

- 6 ☐ **Variable Length** Sequence(s) _____ contain n's or Xaa's which represented more than one residue.
As per the rules, each n or Xaa can only represent a single residue.
Please present the maximum number of each residue having variable length and indicate in the (ix) feature section that some may be missing.

- 7 ☐ **PatentIn ver. 2.0 "bug"** A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequence(s) _____. Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. This applies primarily to the mandatory <220>-<223> sections for Artificial or Unknown sequences.

- 8 ☐ **Skipped Sequences (OLD RULES)** Sequence(s) _____ missing. If intentional, please use the following format for each skipped sequence:
(2) INFORMATION FOR SEQ ID NO:X:
(i) SEQUENCE CHARACTERISTICS:(Do not insert any headings under "SEQUENCE CHARACTERISTICS")
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:X:
This sequence is intentionally skipped

Please also adjust the "(iii) NUMBER OF SEQUENCES:" response to include the skipped sequence(s).

- 9 ☐ **Skipped Sequences (NEW RULES)** Sequence(s) _____ missing. If intentional, please use the following format for each skipped sequence.
<210> sequence id number
<400> sequence id number
000

- 10 ☐ **Use of n's or Xaa's (NEW RULES)** Use of n's and/or Xaa's have been detected in the Sequence Listing.
Use of <220> to <223> is MANDATORY if n's or Xaa's are present.
In <220> to <223> section, please explain location of n or Xaa, and, which residue n or Xaa represents.

- 11 ☐ **Use of "Artificial" (NEW RULES)** Use of "Artificial" only as "<213> Organism" response is incomplete, per 1.823(b) of New Sequence Rules.
Valid response is Artificial Sequence.

- 12 ☐ **Use of <220>Feature (NEW RULES)** Sequence(s) _____ are missing the <220>Feature and associated headings.
Use of <220> to <223> is MANDATORY if <213>ORGANISM is "Artificial Sequence" or "Unknown"
Please explain source of genetic material in <220> to <223> section.
(See "Federal Register," 6/01/98, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of new Rules)

- 13 ☐ **PatentIn ver. 2.0 "bug"** Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing).
Instead, please use "File Manager" or any other means to copy file to floppy disk.

PCT

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/807,802

DATE: 05/01/2001
TIME: 11:39:02

Input Set : A:\sequencelist.txt
Output Set: N:\CRF3\05012001\I807802.raw

Does Not Comply
Corrected Diskette Needed
See pp. 2-8

4 <110> APPLICANT: Wilson, James M.
5 Xiao, Weidong
6 The Trustees of the University of Pennsylvania
8 <120> TITLE OF INVENTION: Adeno-Associated Virus Serotype I Nucleic Acid
9 Sequences, Vectors and Host Cells Containing Same
11 <130> FILE REFERENCE: GNVPN.031PCT
C--> 13 <140> CURRENT APPLICATION NUMBER: US/09/807,802
C--> 14 <141> CURRENT FILING DATE: 2001-04-16
16 <150> PRIOR APPLICATION NUMBER: 60/107,114
17 <151> PRIOR FILING DATE: 1998-11-05
19 <160> NUMBER OF SEQ ID NOS: 20
21 <170> SOFTWARE: PatentIn Ver. 2.0

ERRORED SEQUENCES

23 <210> SEQ ID NO: 1
24 <211> LENGTH: 4718
25 <212> TYPE: DNA
26 <213> ORGANISM: AAV-1
28 <220> FEATURE:
29 <221> NAME/KEY: CDS
30 <222> LOCATION: (335)..(2206)
32 <220> FEATURE:
33 <221> NAME/KEY: CDS
34 <222> LOCATION: (2223)..(4430)
36 <400> SEQUENCE: 1
37 ttgcccactc cctctctgcg cgctcgctcg ctcggtgggg cctgcggacc aaaggtccgc 60
40 agacggcaga gctctgctct gccggcccca ccgagcgagc gagcgcgag agagggagtg 120
43 ggcaactcca tcaactagggg taatcgcgaa gcgcctccca cgctgccgcg tcagcgctga 180
46 cgtaaattac gtcataagggg agtggtcctg tattagctgt cacgtgagtg cttttgcgac 240
49 attttgcgac accacgtggc catttagggg atatatggcc gagtgagcga gcaggatctc 300
52 cattttgacc gcgaaatttg aacgagcagc agcc atg ccg ggc ttc tac gag atc 355
54 Met Pro Gly Phe Tyr Glu Ile
55 1 5
57 gtg atc aag gtg ccg agc gac ctg gac gag cac ctg ccg ggc att tct 403
59 Val Ile Lys Val Pro Ser Asp Leu Asp Glu His Leu Pro Gly Ile Ser
60 10 15 20
62 gac tcg ttt gtg agc tgg gtg gcc gag aag gaa tgg gag ctg ccc ccg 451
64 Asp Ser Phe Val Ser Trp Val Ala Glu Lys Glu Trp Glu Leu Pro Pro
65 25 30 35
67 gat tct gac atg gat ctg aat ctg att gag cag gca ccc ctg acc gtg 499
69 Asp Ser Asp Met Asp Leu Asn Leu Ile Glu Gln Ala Pro Leu Thr Val
70 40 45 50 55
72 gcc gag aag ctg cag cgc gac ttc ctg gtc caa tgg cgc cgc gtg agt 547
74 Ala Glu Lys Leu Gln Arg Asp Phe Leu Val Gln Trp Arg Arg Val Ser
75 60 65 70

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/807,802

DATE: 05/01/2001
TIME: 11:39:02

Input Set : A:\sequencelist.txt
Output Set: N:\CRF3\05012001\I807802.raw

```

77 aag gcc ccg gag gcc ctc ttc ttt gtt cag ttc gag aag ggc gag tcc 595
79 Lys Ala Pro Glu Ala Leu Phe Phe Val Gln Phe Glu Lys Gly Glu Ser
80      75      80      85
82 tac ttc cac ctc cat att ctg gtg gag acc acg ggg gtc aaa tcc atg 643
84 Tyr Phe His Leu His Ile Leu Val Glu Thr Thr Gly Val Lys Ser Met
85      90      95      100
87 gtg ctg ggc cgc ttc ctg agt cag att agg gac aag ctg gtg cag acc 691
89 Val Leu Gly Arg Phe Leu Ser Gln Ile Arg Asp Lys Leu Val Gln Thr
90      105      110      115
92 atc tac cgc ggc atc gag ccg acc ctg ccc aac tgg ttc gcg gtg acc 739
94 Ile Tyr Arg Gly Ile Glu Pro Thr Leu Pro Asn Trp Phe Ala Val Thr
95 120      125      130      135
97 aag acg cgt aat ggc gcc gga ggg ggg aac aag gtg gtg gac gag tgc 787
99 Lys Thr Arg Asn Gly Ala Gly Gly Gly Asn Lys Val Val Asp Glu Cys
100      140      145      150
102 tac atc ccc aac tac ctc ctg ccc aag act cag ccc gag ctg cag tgg 835
104 Tyr Ile Pro Asn Tyr Leu Leu Pro Lys Thr Gln Pro Glu Leu Gln Trp
105      155      160      165
107 gcg tgg act aac atg gag gag tat ata agc gcc tgt ttg aac ctg gcc 883
109 Ala Trp Thr Asn Met Glu Glu Tyr Ile Ser Ala Cys Leu Asn Leu Ala
110      170      175      180
112 gag cgc aaa cgg ctc gtg gcg cag cac ctg acc cac gtc agc cag acc 931
114 Glu Arg Lys Arg Leu Val Ala Gln His Leu Thr His Val Ser Gln Thr
115      185      190      195
117 cag gag cag aac aag gag aat ctg aac ccc aat tct gac gcg cct gtc 979
119 Gln Glu Gln Asn Lys Glu Asn Leu Asn Pro Asn Ser Asp Ala Pro Val
120 200      205      210      215
E--> 122 atc cgg tca aaa acc tcc gcg cgc tac atg gag ctg gtc ggg tgg ctg
123 1027
124 Ile Arg Ser Lys Thr Ser Ala Arg Tyr Met Glu Leu Val Gly Trp Leu
125      220      225      230
E--> 127 gtg gac cgg ggc atc acc tcc gag aag cag tgg atc cag gag gac cag
128 1075
129 Val Asp Arg Gly Ile Thr Ser Glu Lys Gln Trp Ile Gln Glu Asp Gln
130      235      240      245
E--> 132 gcc tcg tac atc tcc ttc aac gcc gct tcc aac tcg cgg tcc cag atc
133 1123
134 Ala Ser Tyr Ile Ser Phe Asn Ala Ala Ser Asn Ser Arg Ser Gln Ile
135      250      255      260
E--> 137 aag gcc gct ctg gac aat gcc ggc aag atc atg gcg ctg acc aaa tcc
138 1171
139 Lys Ala Ala Leu Asp Asn Ala Gly Lys Ile Met Ala Leu Thr Lys Ser
140      265      270      275
E--> 142 gcg ccc gac tac ctg gta ggc ccc gct ccg ccc gcg gac att aaa acc
143 1219
144 Ala Pro Asp Tyr Leu Val Gly Pro Ala Pro Pro Ala Asp Ile Lys Thr
145 280      285      290      295
E--> 147 aac cgc atc tac cgc atc ctg gag ctg aac ggc tac gaa cct gcc tac
148 1267

```

End of line
nucleic number
has "wrapped"
down to the
next line. See
#1 on the
Error Summary
sheet.

RAW SEQUENCE LISTING
PATENT APPLICATION: .US/09/807,802

DATE: 05/01/2001
TIME: 11:39:02

Input Set : A:\sequencelist.txt
Output Set: N:\CRF3\05012001\I807802.raw

```

149 Asn Arg Ile Tyr Arg Ile Leu Glu Leu Asn Gly Tyr Glu Pro Ala Tyr
150                               300                               305                               310
E--> 152 gcc ggc tcc gtc ttt ctc ggc tgg gcc cag aaa agg ttc ggg aag cgc
153 (1315)
154 Ala Gly Ser Val Phe Leu Gly Trp Ala Gln Lys Arg Phe Gly Lys Arg
155                               315                               320                               325
E--> 157 aac acc atc tgg ctg ttt ggg ccg gcc acc acg ggc aag acc aac atc
158 (1363)
159 Asn Thr Ile Trp Leu Phe Gly Pro Ala Thr Thr Gly Lys Thr Asn Ile
160                               330                               335                               340
E--> 162 gcg gaa gcc atc gcc cac gcc gtg ccc ttc tac ggc tgc gtc aac tgg
163 (1411)
164 Ala Glu Ala Ile Ala His Ala Val Pro Phe Tyr Gly Cys Val Asn Trp
165                               345                               350                               355
E--> 167 acc aat gag aac ttt ccc ttc aat gat tgc gtc gac aag atg gtg atc
168 (1459)
169 Thr Asn Glu Asn Phe Pro Phe Asn Asp Cys Val Asp Lys Met Val Ile
170 360                               365                               370                               375
E--> 172 tgg tgg gag gag ggc aag atg acg gcc aag gtc gtg gag tcc gcc aag
173 (1507)
174 Trp Trp Glu Glu Gly Lys Met Thr Ala Lys Val Val Glu Ser Ala Lys
175                               380                               385                               390
E--> 177 gcc att ctc ggc ggc agc aag gtg cgc gtg gac caa aag tgc aag tcg
178 (1555)
179 Ala Ile Leu Gly Gly Ser Lys Val Arg Val Asp Gln Lys Cys Lys Ser
180                               395                               400                               405
E--> 182 tcc gcc cag atc gac ccc acc ccc gtg atc gtc acc tcc aac acc aac
183 (1603)
184 Ser Ala Gln Ile Asp Pro Thr Pro Val Ile Val Thr Ser Asn Thr Asn
185                               410                               415                               420
E--> 187 atg tgc gcc gtg att gac ggg aac agc acc acc ttc gag cac cag cag
188 (1651)
189 Met Cys Ala Val Ile Asp Gly Asn Ser Thr Thr Phe Glu His Gln Gln
190                               425                               430                               435
E--> 192 ccg ttg cag gac cgg atg ttc aaa ttt gaa ctc acc cgc cgt ctg gag
193 (1699)
194 Pro Leu Gln Asp Arg Met Phe Lys Phe Glu Leu Thr Arg Arg Leu Glu
195 440                               445                               450                               455
E--> 197 cat gac ttt ggc aag gtg aca aag cag gaa gtc aaa gag ttc ttc cgc
198 (1747)
199 His Asp Phe Gly Lys Val Thr Lys Gln Glu Val Lys Glu Phe Phe Arg
200                               460                               465                               470
E--> 202 tgg gcg cag gat cac gtg acc gag gtg gcg cat gag ttc tac gtc aga
203 (1795)
204 Trp Ala Gln Asp His Val Thr Glu Val Ala His Glu Phe Tyr Val Arg
205                               475                               480                               485
E--> 207 aag ggt gga gcc aac aaa aga ccc gcc ccc gat gac gcg gat aaa agc
208 (1843)
209 Lys Gly Gly Ala Asn Lys Arg Pro Ala Pro Asp Asp Ala Asp Lys Ser

```

"wrapped"
nucleic
numbers.
see p-2

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/807,802

DATE: 05/01/2001

TIME: 11:39:02

Input Set : A:\sequencelist.txt

Output Set: N:\CRF3\05012001\I807802.raw

```

      210      490      495      500
E--> 212 gag ccc aag cgg gcc tgc ccc tca gtc gcg gat cca tcg acg tca gac
      213 1891
      214 Glu Pro Lys Arg Ala Cys Pro Ser Val Ala Asp Pro Ser Thr Ser Asp
      215      505      510      515
E--> 217 gcg gaa gga gct ccg gtg gac ttt gcc gac agg tac caa aac aaa tgt
      218 1939
      219 Ala Glu Gly Ala Pro Val Asp Phe Ala Asp Arg Tyr Gln Asn Lys Cys
      220 520      525      530      535
E--> 222 tct cgt cac gcg ggc atg ctt cag atg ctg ttt ccc tgc aag aca tgc
      223 1987
      224 Ser Arg His Ala Gly Met Leu Gln Met Leu Phe Pro Cys Lys Thr Cys
      225      540      545      550
E--> 227 gag aga atg aat cag aat ttc aac att tgc ttc acg cac ggg acg aga
      228 2035
      229 Glu Arg Met Asn Gln Asn Phe Asn Ile Cys Phe Thr His Gly Thr Arg
      230      555      560      565
E--> 232 gac tgt tca gag tgc ttc ccc ggc gtg tca gaa tct caa ccg gtc gtc
      233 2083
      234 Asp Cys Ser Glu Cys Phe Pro Gly Val Ser Glu Ser Gln Pro Val Val
      235      570      575      580
E--> 237 aga aag agg acg tat cgg aaa ctc tgt gcc att cat cat ctg ctg ggg
      238 2131
      239 Arg Lys Arg Thr Tyr Arg Lys Leu Cys Ala Ile His His Leu Leu Gly
      240      585      590      595
E--> 242 cgg gct ccc gag att gct tgc tgc gcc tgc gat ctg gtc aac gtg gac
      243 2179
      244 Arg Ala Pro Glu Ile Ala Cys Ser Ala Cys Asp Leu Val Asn Val Asp
      245 600      605      610      615
E--> 247 ctg gat gac tgt gtt tct gag caa taa atgacttaaa ccaggt atg gct gcc
      248 2231
      249 Leu Asp Asp Cys Val Ser Glu Gln      Met Ala Ala
W--> 250      620      625
E--> 252 gat ggt tat ctt cca gat tgg ctc gag gac aac ctc tct gag ggc att
      253 2279
      254 Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser Glu Gly Ile
W--> 255      630      635      640
E--> 257 cgc gag tgg tgg gac ttg aaa cct gga gcc ccg aag ccc aaa gcc aac
      258 2327
      259 Arg Glu Trp Trp Asp Leu Lys Pro Gly Ala Pro Lys Pro Lys Ala Asn
W--> 260      645      650      655
E--> 262 cag caa aag cag gac gac ggc cgg ggt ctg gtg ctt cct ggc tac aag
      263 2375
      264 Gln Gln Lys Gln Asp Asp Gly Arg Gly Leu Val Leu Pro Gly Tyr Lys
W--> 265 660      665      670      675
E--> 267 tac ctc gga ccc ttc aac gga ctc gac aag ggg gag ccc gtc aac gcg
      268 2423
      269 Tyr Leu Gly Pro Phe Asn Gly Leu Asp Lys Gly Glu Pro Val Asn Ala
W--> 270      680      685      690

```

Same

RAW SEQUENCE LISTING
PATENT APPLICATION: . US/09/807,802

DATE: 05/01/2001
TIME: 11:39:02

Input Set : A:\sequencelist.txt
Output Set: N:\CRF3\05012001\I807802.raw

```

E--> 272 gcg gac gca gcg gcc ctc gag cac gac aag gcc tac gac cag cag ctc
      273 (2471)
      274 Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp Gln Gln Leu
W--> 275          695          700          705
E--> 277 aaa gcg ggt gac aat ccg tac ctg cgg tat aac cac gcc gac gcc gag
      278 (2519)
      279 Lys Ala Gly Asp Asn Pro Tyr Leu Arg Tyr Asn His Ala Asp Ala Glu
W--> 280          710          715          720
E--> 282 ttt cag gag cgt ctg caa gaa gat acg tct ttt ggg ggc aac ctc ggg
      283 (2567)
      284 Phe Gln Glu Arg Leu Gln Glu Asp Thr Ser Phe Gly Gly Asn Leu Gly
W--> 285          725          730          735
E--> 287 cga gca gtc ttc cag gcc aag aag cgg gtt ctc gaa cct ctc ggt ctg
      288 (2615)
      289 Arg Ala Val Phe Gln Ala Lys Lys Arg Val Leu Glu Pro Leu Gly Leu
W--> 290 740          745          750          755
E--> 292 gtt gag gaa ggc gct aag acg gct cct gga aag aaa cgt ccg gta gag
      293 (2663)
      294 Val Glu Glu Gly Ala Lys Thr Ala Pro Gly Lys Lys Arg Pro Val Glu
W--> 295          760          765          770
E--> 297 cag tcg cca caa gag cca gac tcc tcc tcg ggc atc ggc aag aca ggc
      298 (2711)
      299 Gln Ser Pro Gln Glu Pro Asp Ser Ser Ser Gly Ile Gly Lys Thr Gly
W--> 300          775          780          785
E--> 302 cag cag ccc gct aaa aag aga ctc aat ttt ggt cag act ggc gac tca
      303 (2759)
      304 Gln Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr Gly Asp Ser
W--> 305          790          795          800
E--> 307 gag tca gtc ccc gat cca caa cct ctc gga gaa cct cca gca acc ccc
      308 (2807)
      309 Glu Ser Val Pro Asp Pro Gln Pro Leu Gly Glu Pro Pro Ala Thr Pro
W--> 310          805          810          815
E--> 312 gct gct gtg gga cct act aca atg gct tca ggc ggt ggc gca cca atg
      313 (2855)
      314 Ala Ala Val Gly Pro Thr Thr Met Ala Ser Gly Gly Gly Ala Pro Met
W--> 315 820          825          830          835
E--> 317 gca gac aat aac gaa ggc gcc gac gga gtg ggt aat gcc tca gga aat
      318 (2903)
      319 Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Asn Ala Ser Gly Asn
W--> 320          840          845          850
E--> 322 tgg cat tgc gat tcc aca tgg ctg ggc gac aga gtc atc acc acc agc
      323 (2951)
      324 Trp His Cys Asp Ser Thr Trp Leu Gly Asp Arg Val Ile Thr Thr Ser
W--> 325          855          860          865
E--> 327 acc cgc acc tgg gcc ttg ccc acc tac aat aac cac ctc tac aag caa
      328 (2999)
      329 Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu Tyr Lys Gln
W--> 330          870          875          880
E--> 332 atc tcc agt gct tca acg ggg gcc agc aac gac aac cac tac ttc ggc

```

Same

RAW SEQUENCE LISTING
PATENT APPLICATION: .US/09/807,802

DATE: 05/01/2001
TIME: 11:39:02

Input Set : A:\sequencelist.txt
Output Set: N:\CRF3\05012001\I807802.raw

```

333 3047
334 Ile Ser Ser Ala Ser Thr Gly Ala Ser Asn Asp Asn His Tyr Phe Gly
W--> 335      885      890      895
E--> 337 tac agc acc ccc tgg ggg tat ttt gat ttc aac aga ttc cac tgc cac
338 3095
339 Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg Phe His Cys His
W--> 340 900      905      910      915
E--> 342 ttt tca cca cgt gac tgg cag cga ctc atc aac aac aat tgg gga ttc
343 3143
344 Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn Asn Trp Gly Phe
W--> 345      920      925      930
E--> 347 cgg ccc aag aga ctc aac ttc aaa ctc ttc aac atc caa gtc aag gag
348 3191
349 Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile Gln Val Lys Glu
W--> 350      935      940      945
E--> 352 gtc acg acg aat gat ggc gtc aca acc atc gct aat aac ctt acc agc
353 3239
354 Val Thr Thr Asn Asp Gly Val Thr Thr Ile Ala Asn Asn Leu Thr Ser
W--> 355      950      955      960
E--> 357 acg gtt caa gtc ttc tcg gac tcg gag tac cag ctt ccg tac gtc ctc
358 3287
359 Thr Val Gln Val Phe Ser Asp Ser Glu Tyr Gln Leu Pro Tyr Val Leu
W--> 360      965      970      975
E--> 362 ggc tct gcg cac cag ggc tgc ctc cct ccg ttc ccg gcg gac gtg ttc
363 3335
364 Gly Ser Ala His Gln Gly Cys Leu Pro Pro Phe Pro Ala Asp Val Phe
W--> 365 980      985      990      995
E--> 367 atg att ccg caa tac ggc tac ctg acg ctc aac aat ggc agc caa gcc
368 3383
369 Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asn Gly Ser Gln Ala
W--> 370      1000      1005      1010
E--> 372 gtg gga cgt tca tcc ttt tac tgc ctg gaa tat ttc cct tct cag atg
373 3431
374 Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe Pro Ser Gln Met
W--> 375      1015      1020      1025
E--> 377 ctg aga acg ggc aac aac ttt acc ttc agc tac acc ttt gag gaa gtg
378 3479
379 Leu Arg Thr Gly Asn Asn Phe Thr Phe Ser Tyr Thr Phe Glu Glu Val
W--> 380      1030      1035      1040
E--> 382 cct ttc cac agc agc tac gcg cac agc cag agc ctg gac cgg ctg atg
383 3527
384 Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu Asp Arg Leu Met
W--> 385      1045      1050      1055
E--> 387 aat cct ctc atc gac caa tac ctg tat tac ctg aac aga act caa aat
388 3575
389 Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Asn Arg Thr Gln Asn
W--> 390 1060      1065      1070      1075
E--> 392 cag tcc gga agt gcc caa aac aag gac ttg ctg ttt agc cgt ggg tct
393 3623

```

Same

RAW SEQUENCE LISTING
PATENT APPLICATION: . US/09/807,802

DATE: 05/01/2001
TIME: 11:39:02

Input Set : A:\sequencelist.txt
Output Set: N:\CRF3\05012001\I807802.raw

```

394 Gln Ser Gly Ser Ala Gln Asn Lys Asp Leu Leu Phe Ser Arg Gly Ser
W--> 395          1080          1085          1090
E--> 397 cea-gct ggc atg tct gtt cag ccc aaa aac tgg cta cct gga ccc tgt
398 3671
399 Pro Ala Gly Met Ser Val Gln Pro Lys Asn Trp Leu Pro Gly Pro Cys
W--> 400          1095          1100          1105
E--> 402 tat cgg cag cag cgc gtt tct aaa aca aaa aca gac aac aac aac agc
403 3719
404 Tyr Arg Gln Gln Arg Val Ser Lys Thr Lys Thr Asp Asn Asn Asn Ser
W--> 405          1110          1115          1120
E--> 407 aat ttt acc tgg act ggt gct tca aaa tat aac ctc aat ggg cgt gaa
408 3767
409 Asn Phe Thr Trp Thr Gly Ala Ser Lys Tyr Asn Leu Asn Gly Arg Glu
W--> 410          1125          1130          1135
E--> 412 tcc atc atc aac cct ggc act gct atg gcc tca cac aaa gac gac gaa
413 3815
414 Ser Ile Ile Asn Pro Gly Thr Ala Met Ala Ser His Lys Asp Asp Glu
W--> 415 1140          1145          1150          1155
E--> 417 gac aag ttc ttt ccc atg agc ggt gtc atg att ttt gga aaa gag agc
418 8863
419 Asp Lys Phe Phe Pro Met Ser Gly Val Met Ile Phe Gly Lys Glu Ser
W--> 420          1160          1165          1170
E--> 422 gcc gga gct tca aac act gca ttg gac aat gtc atg att aca gac gaa
423 3911
424 Ala Gly Ala Ser Asn Thr Ala Leu Asp Asn Val Met Ile Thr Asp Glu
W--> 425          1175          1180          1185
E--> 427 gag gaa att aaa gcc act aac cct gtg gcc acc gaa aga ttt ggg acc
428 3959
429 Glu Glu Ile Lys Ala Thr Asn Pro Val Ala Thr Glu Arg Phe Gly Thr
W--> 430          1190          1195          1200
E--> 432 gtg gca gtc aat ttc cag agc agc agc aca gac cct gcg acc gga gat
433 4007
434 Val Ala Val Asn Phe Gln Ser Ser Ser Thr Asp Pro Ala Thr Gly Asp
W--> 435          1205          1210          1215
E--> 437 gtg cat gct atg gga gca tta cct ggc atg gtg tgg caa gat aga gac
438 4055
439 Val His Ala Met Gly Ala Leu Pro Gly Met Val Trp Gln Asp Arg Asp
W--> 440 1220          1225          1230          1235
E--> 442 gtg tac ctg cag ggt ccc att tgg gcc aaa att cct cac aca gat gga
443 4103
444 Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His Thr Asp Gly
W--> 445          1240          1245          1250
E--> 447 cac ttt cac ccg tct cct ctt atg ggc ggc ttt gga ctc aag aac ccg
448 4151
449 His Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Leu Lys Asn Pro
W--> 450          1255          1260          1265
E--> 452 cct cct cag atc ctc atc aaa aac acg cct gtt cct gcg aat cct ccg
453 4199
454 Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala Asn Pro Pro

```

Same

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/807,802

DATE: 05/01/2001
TIME: 11:39:02

Input Set : A:\sequencelist.txt
Output Set: N:\CRF3\05012001\I807802.raw

```

W--> 455      1270      1275      1280
E--> 457 gcg gag ttt tca gct aca aag ttt gct tca ttc atc acc caa tac tcc
458 4247
459 Ala Glu Phe Ser Ala Thr Lys Phe Ala Ser Phe Ile Thr Gln Tyr Ser
W--> 460      1285      1290      1295
E--> 462 aca gga caa gtg agt gtg gaa att gaa tgg gag ctg cag aaa gaa aac
463 4295
464 Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln Lys Glu Asn
W--> 465 1300      1305      1310      1315
E--> 467 agc aag cgc tgg aat ccc gaa gtg cag tac aca tcc aat tat gca aaa
468 4343
469 Ser Lys Arg Trp Asn Pro Glu Val Gln Tyr Thr Ser Asn Tyr Ala Lys
W--> 470      1320      1325      1330
E--> 472 tct gcc aac gtt gat ttt act gtg gac aac aat gga ctt tat act gag
473 4391
474 Ser Ala Asn Val Asp Phe Thr Val Asp Asn Asn Gly Leu Tyr Thr Glu
W--> 475      1335      1340      1345
E--> 477 pct cgc ccc att ggc acc cgt tac ctt acc cgt ccc ctg taattacgtg
478 4440
479 Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Pro Leu
W--> 480      1350      1355      1360
E--> 482 ttaatcaata aaccggttga ttcgtttcag ttgaactttg gtctcctgtc cttcttatct
483 4500
485 tateggttac catggttata gcttacacat taactgcttg gttgcgcttc gcgataaaag
486 4560
488 attacgtca tcgggttacc cctagtgatg gagttgccca ctccctctct gcgcgctcgc
489 4620
E--> 491 tcgctcggtg gggcctgcgg accaaaggtc cgcagacggc agagctctgc tctgccggcc
492 4680
E--> 494 ccaccgagcg agcgagcgcg cagagaggga gtgggcaa
495 4718
767 <210> SEQ ID NO: 4
768 <211> LENGTH: 1872
769 <212> TYPE: DNA
770 <213> ORGANISM: AAV-1
772 <220> FEATURE:
773 <221> NAME/KEY: CDS
774 <222> LOCATION: (1)..(1869)
776 <400> SEQUENCE: 4
777 atg ccg ggc ttc tac gag atc gtg atc aag gtg ccg agc gac ctg gac 48
779 Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
780 1 5 10 15
782 gag cac ctg ccg ggc att tct gac tcg ttt gtg agc tgg gtg gcc gag 96
784 Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Ser Trp Val Ala Glu
785 20 25 30
787 aag gaa tgg gag ctg ccc ccg gat tct gac atg gat ctg aat ctg att 144
789 Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
790 35 40 45
792 gag cag gca ccc ctg acc gtg gcc gag aag ctg cag cgc gac ttc ctg 192

```

Same

Note

→ The types of errors shown exist throughout the Sequence Listing. Please check subsequent sequences for similar errors.

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/807,802

DATE: 05/01/2001

TIME: 11:39:03

Input Set : A:\sequencelist.txt

Output Set: N:\CRF3\05012001\I807802.raw

L:13 M:270 C: Current Application Number differs, Replaced Application Number
L:14 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:122 M:254 E: No. of Bases conflict, LENGTH:Input:0 Counted:1027 SEQ:1
M:254 Repeated in SeqNo=1
L:250 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:255 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:260 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:265 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:270 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:275 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:280 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:285 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:290 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:295 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:300 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:305 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:310 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:315 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:320 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:325 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:330 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:335 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:340 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:345 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:350 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:355 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:360 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:365 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:370 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:375 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:380 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:385 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:390 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:395 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:400 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:405 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:410 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:415 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:420 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:425 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:430 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:435 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:440 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:445 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:450 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:455 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:460 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:465 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1

VERIFICATION SUMMARY

PATENT APPLICATION: 'US/09/807,802

DATE: 05/01/2001

TIME: 11:39:03

Input Set : A:\sequencelist.txt

Output Set: N:\CRF3\05012001\I807802.raw

L:470 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:475 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:480 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:877 M:254 E: No. of Bases conflict, LENGTH:Input:0 Counted:1008 SEQ:4
M:254 Repeated in SeqNo=4
L:1207 M:254 E: No. of Bases conflict, LENGTH:Input:0 Counted:1008 SEQ:6
M:254 Repeated in SeqNo=6
L:1505 M:254 E: No. of Bases conflict, LENGTH:Input:0 Counted:1008 SEQ:8
M:254 Repeated in SeqNo=8
L:1913 M:254 E: No. of Bases conflict, LENGTH:Input:0 Counted:1008 SEQ:12
M:254 Repeated in SeqNo=12
L:2302 M:254 E: No. of Bases conflict, LENGTH:Input:0 Counted:1008 SEQ:14
M:254 Repeated in SeqNo=14
L:2624 M:254 E: No. of Bases conflict, LENGTH:Input:0 Counted:1008 SEQ:16
M:254 Repeated in SeqNo=16
L:2858 M:254 E: No. of Bases conflict, LENGTH:Input:0 Counted:1020 SEQ:18
M:254 Repeated in SeqNo=18
L:3102 M:254 E: No. of Bases conflict, LENGTH:Input:0 Counted:1020 SEQ:19
M:254 Repeated in SeqNo=19